

When One Code Defines Millions: Sensitivity of Peripheral Artery Disease Cohorts to Operational Definition Choices in Real-World Data

Scott L DuVall PhD,¹ Jared Kamaau BFA,² Aimee Harrison MFA,² Michael Buck PhD,² Craig Parker MD MS,² Allise Kamaau MS,² Aaron Kamaau MD MS MPH²

¹PurpleLab Healthcare Analytics, USA; ²Navidence Inc., USA

So What?

Operational definition choices can produce significant differences in cohort size – driven by a small number of high-impact diagnosis codes rather than overall code list size.

BACKGROUND

Context

Composite phenotypes such as peripheral artery disease (PAD) are commonly defined in real-world research (RWR) using diagnosis code lists derived from prior studies. Published code lists vary widely in size and composition

- Peripheral Artery Disease (PAD): from 52-351 ICD-10-CM codes
- Myocardial Infarction (MI): from 22-25 ICD-10-CM codes
- Stroke: 118-171 ICD-10-CM codes
- Transient Ischemic Attack (TIA): 2-11 ICD-10-CM codes

The Gap / Problem

While variability in these operational definitions is recognized (see companion posters *MSR49* & *PT7*), the population-level impact of individual codes is rarely quantified. It is unclear whether cohort size is driven by the number of codes or by which specific codes are included.

OBJECTIVES

Primary Objective

To assess the sensitivity of PAD cohort size to operational definition choices in real-world data, and to identify the diagnosis codes that most strongly drive cohort inclusion.

Secondary Objectives

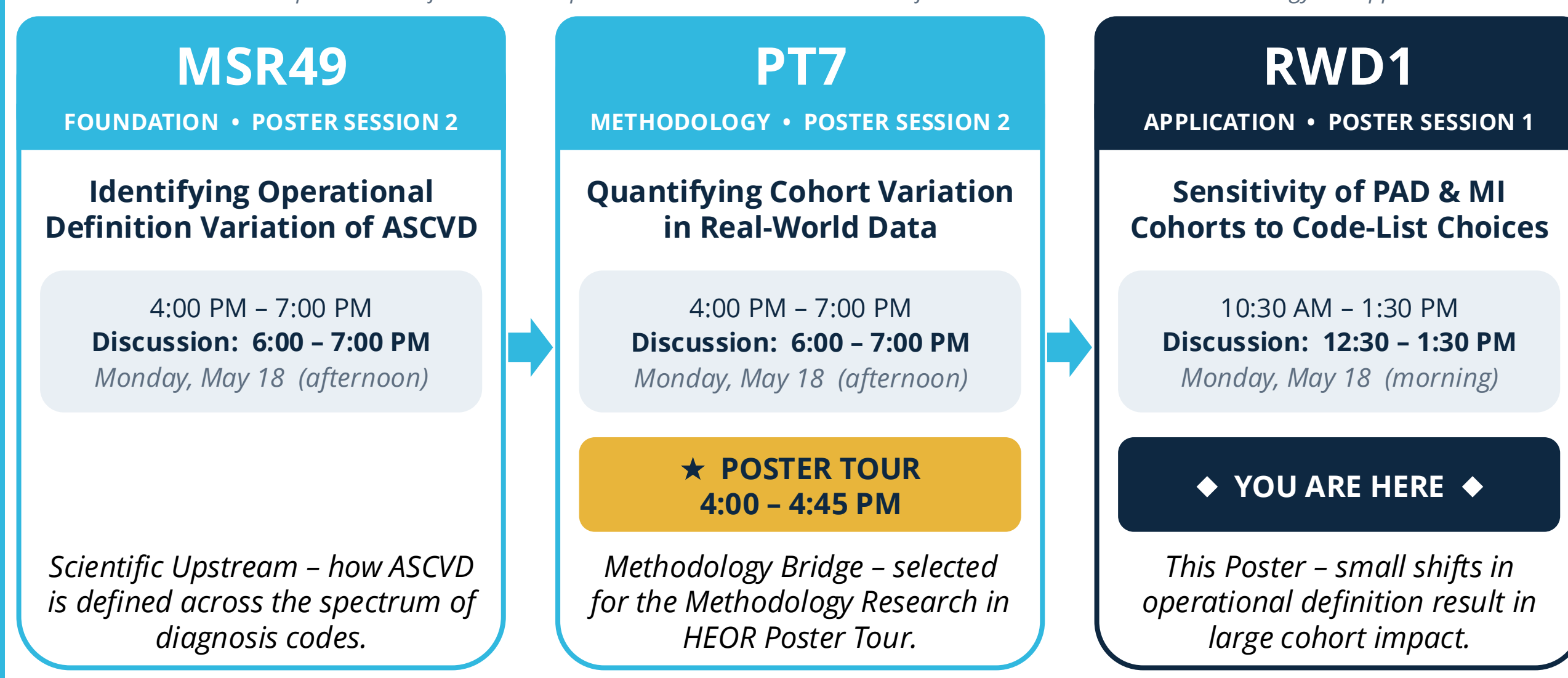
- Quantify code-level contributions to PAD cohort inclusion
- Identify individual high-impact ICD-10-CM codes that disproportionately affect cohort size
- Evaluate generalizability via parallel sensitivity assessments for MI, Stroke and TIA

REFERENCES

- FDA. Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices. 2023.
- Kamaau J, et al. Defining ASCVD in RWR: An Analysis of Conceptual and Operational Definition Variability. ISPOR 2026; Poster MSR49.
- Kamaau A, et al. Impact of Variance in Cohort Phenotype Definitions on RWR: ASCVD Assessment. ISPOR 2026; Poster PT7.
- Mackinnon ES, et al. Increasing Prevalence and Incidence of ASCVD in Adult Patients in Ontario, Canada. CJC Open. 2021;4(2):206-213.
- Dai D, et al. Multimorbidity in ASCVD and Its Associations With Adverse CV Events and Healthcare Costs. J Health Econ Outcomes Res. 2024;11(1):75-85.
- Sajja A, et al. A simplified approach to identification of risk status in patients with ASCVD. Am J Prev Cardiol. 2021;7:100187.
- Barrios V, et al. Treatment patterns and use of healthcare resources of patients with ASCVD. Front Cardiovasc Med. 2022;9:966049.
- Mackinnon ES, et al. Real-World Risk of Recurrent CV Events in ASCVD Patients. Cardiol Ther. 2024;13(1):205-220.

PART OF A 3-POSTER SERIES • ISPOR 2026 • MONDAY MAY 18

The work in this poster stems from two companion studies — see them this afternoon: Foundation • Methodology • Application

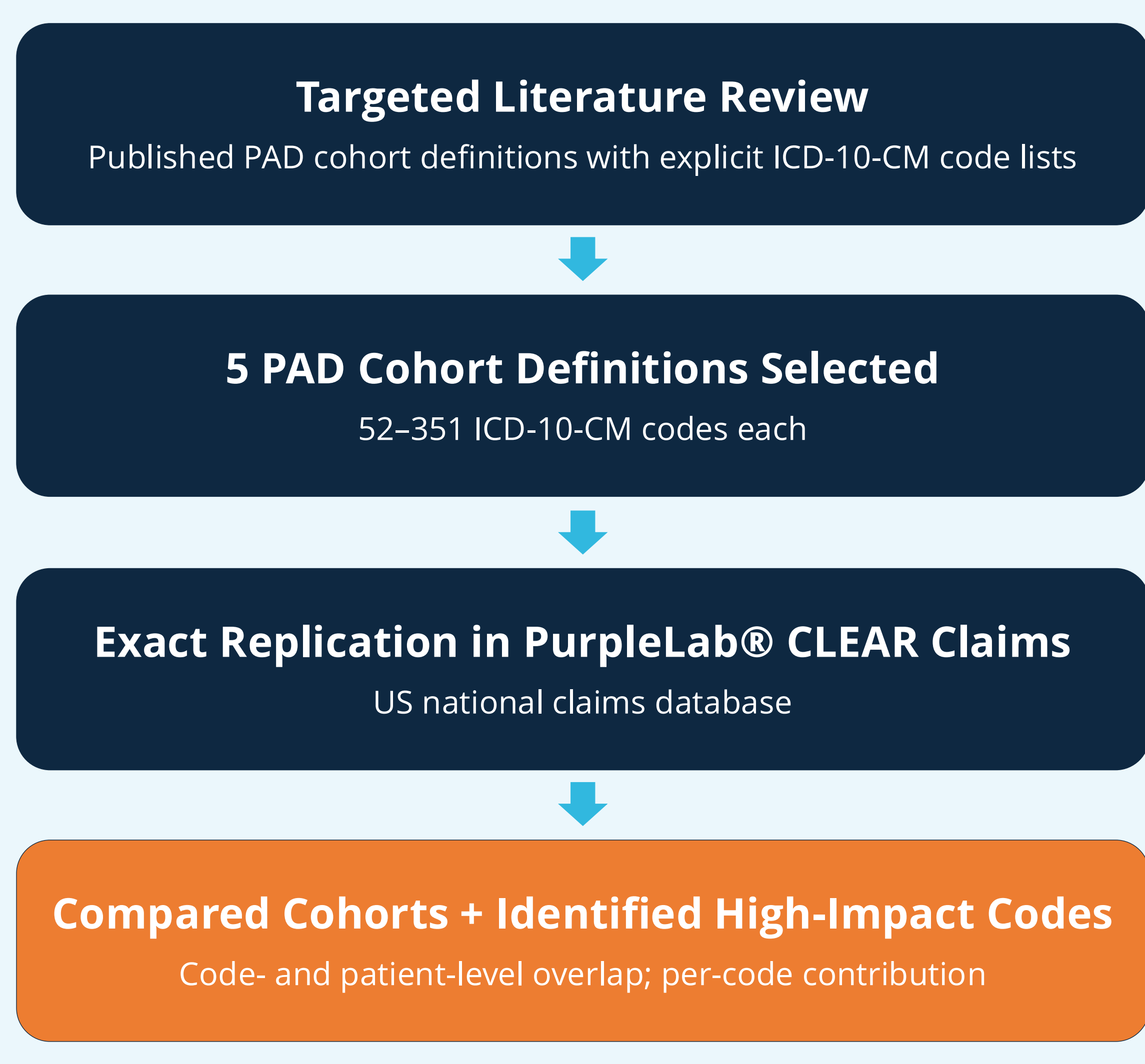


METHODS

Study Design & Data Source

Cross-sectional cohort replication study. Multiple published real-world cohort definitions for PAD, MI and TIA, with explicit ICD-10-CM code lists were identified from the literature, then precisely replicated (matching code lists exactly) in PurpleLab® CLEAR Claims, a US claims database.

Figure 1: PAD Cohort Replication Workflow



Computable Operational Definitions (CODEfs™)

Study elements — exposures, outcomes, and covariates — were defined using Navidence CODEfs™: rigorous, data source-agnostic algorithms that ensure precision and reproducibility across RWR.

RESULTS

5 PAD Cohort Definitions Replicated
52–351 ICD-10-CM Codes per PAD Cohort Definition
5.2–6.9M Patients Identified across the 5 PAD cohorts

Figure 2: PAD Code List Size vs. Cohort Size

~6x more codes → only ~1.3x more patients

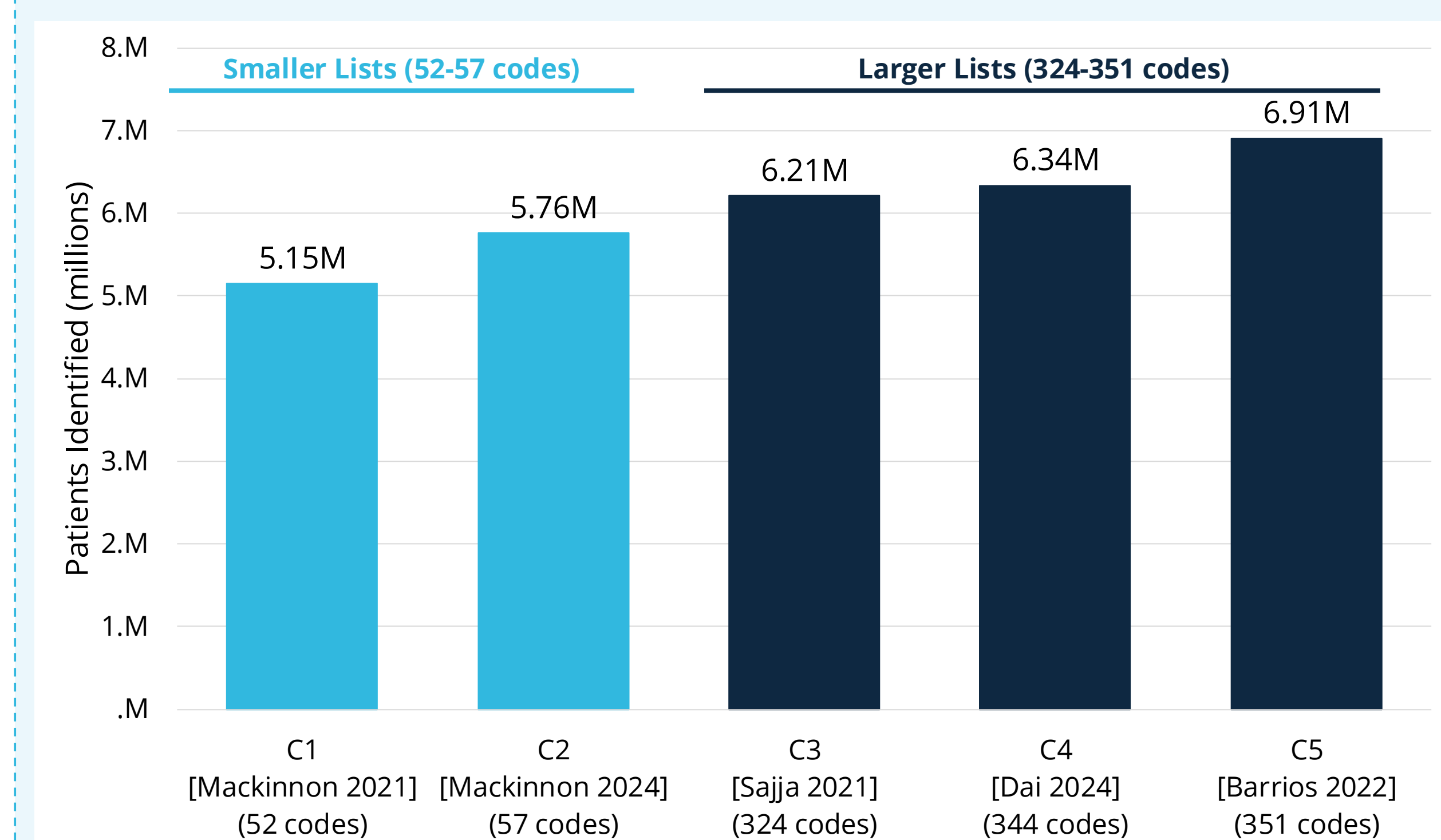
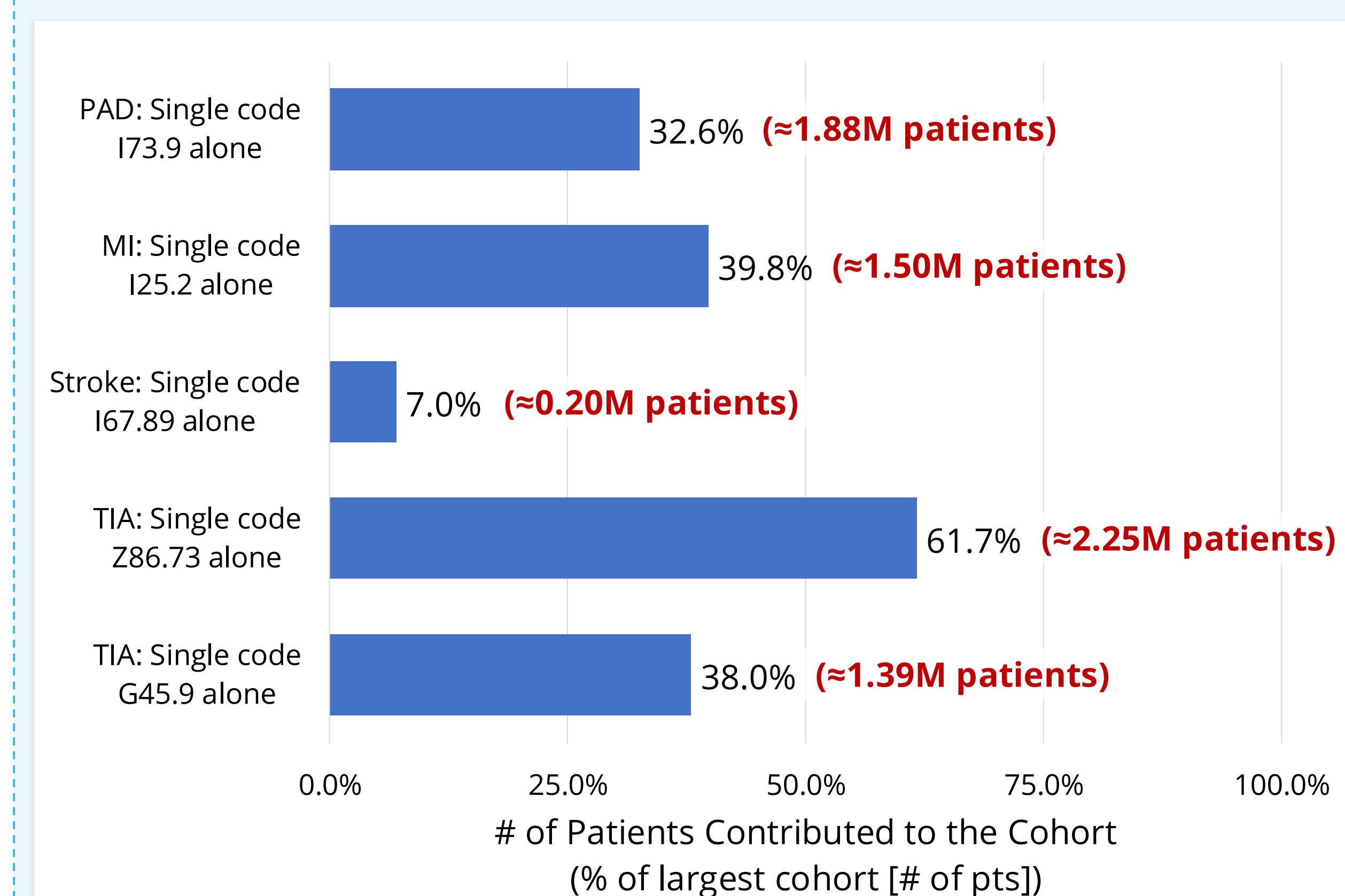


Figure 3: Single-Code Impact

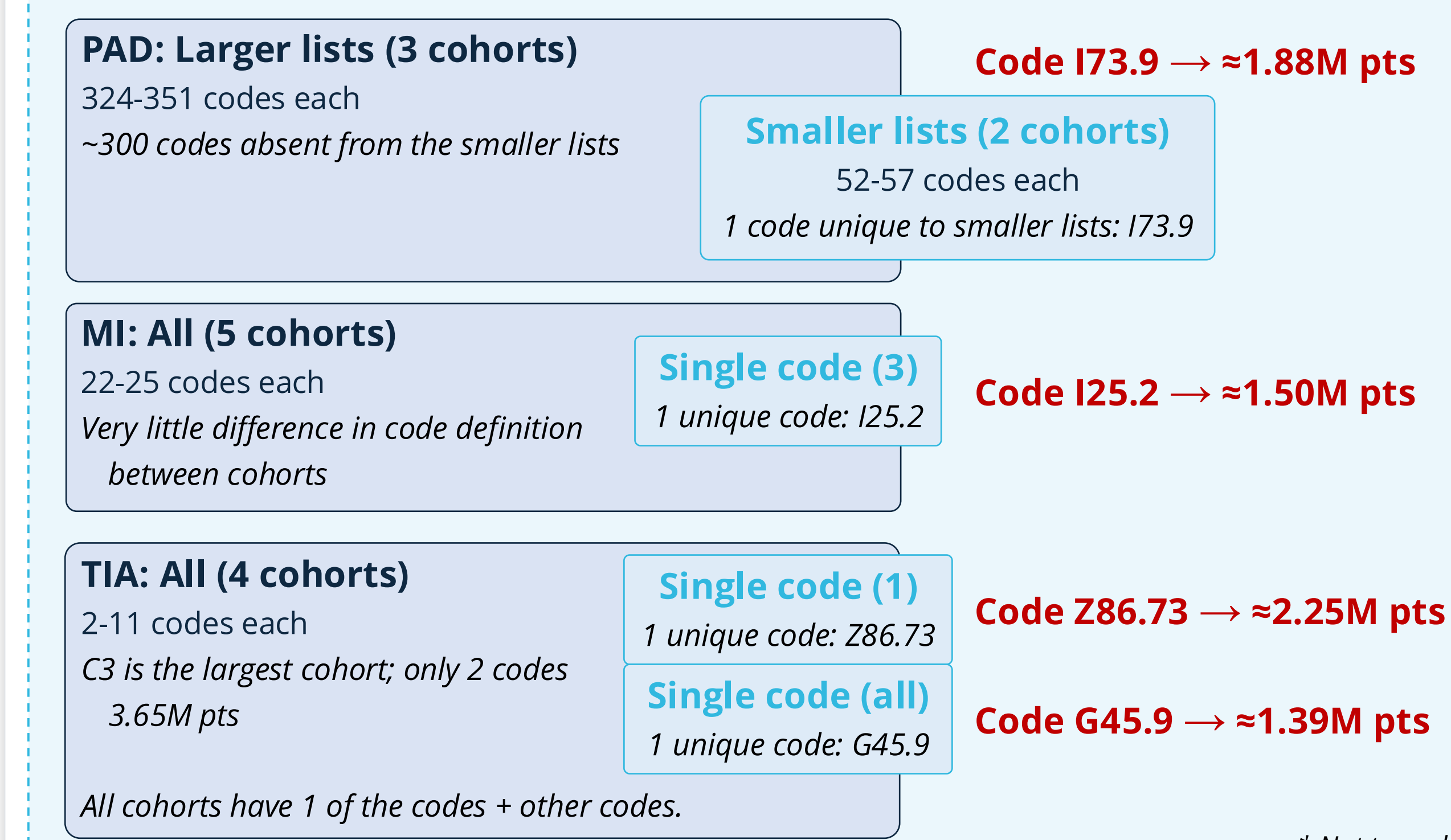
A single ICD-10-CM code can drive cohort inclusion



ICD-10-CM Codes:

- I73.9 - Peripheral vascular disease, unspecified
- I25.2 - Old myocardial infarction
- I67.89 - Other cerebrovascular disease
- Z86.73 - Personal history of transient ischemic attack (TIA), and cerebral infarction without residual deficits
- G45.9 - Transient cerebral ischemic attack, unspecified

Figure 4: Code List & Code Impact Schematic



CONCLUSIONS

Key Findings

- Five PAD cohort definitions (52-351 ICD-10-CM codes) identified 5.15–6.91M patient in PurpleLab® CLEAR Claims.
- Three larger lists (324-351 codes) identified only ~1.3x more patients than the two smaller lists (52, 57 codes) despite ~6x difference in code list size.
- A single code (I73.9 “Peripheral vascular disease, unspecified”) absent from all three larger lists contributed ≈ 1.88M patients to the smaller cohorts.
- Parallel assessments of MI, stroke, and TIE show similar sensitivity.

Clinical Research Implications

- Code list size is not a proxy for phenotype completeness – which codes are included matters more than how many.
- Data-informed phenotype design aligned with study intent is essential for RWE generation.
- Researchers should publish complete code lists and code-level contribution to cohort definition.

LIMITATIONS

Findings are based on PurpleLab® CLEAR Claims; results from other RWD sources may vary.
 Code-level characterizations of other sub-populations planned as next step.